Characterization of non-Gaussianity in gravitational wave detector noise

Takahiro Yamamoto,[1] Kazuhiro Hayama,[2,3] Shuhei Mano,[4] Yousuke Itoh,[5] and Nobuyuki Kanda[1]
[1]*Department of Physics, Osaka City University, 3-3-138, Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan*
[2]*KAGRA Observatory, Institute for Cosmic Ray Research, University of Tokyo, 238 Higashi Mozumi, Kamioka, Hida, Gifu 506-1205, Japan*
[3]*Gravitational Wave Project Office, National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Tokyo 181-8588, Japan*
[4]*The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*
[5]*Research Center for the Early Universe (RESCEU), Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan*
(Received 24 February 2016; published 26 April 2016)

The first detection of a gravitational wave (GW) has been achieved by two detectors of the advanced LIGO. Routine detections of GW events from various GW sources are expected in the coming decades. Although the first signal was statistically significant, we expect to see numerous low signal-to-noise ratio (SNR) events with which we may be able to learn various aspects of the Universe that have yet to be unveiled. On the other hand, instrumental glitches due to nonstationarity and/or a non-Gaussian tail of detector noise distribution prevent us from confidently identifying true but low SNR GW signals out of instrumental noise. Thus, to make the best use of data from GW detectors, it is important to establish a method to safely distinguish true GW signals from false signals due to instrumental noises. For this purpose, we urgently need to understand characteristics of detector noises, since the nonstationarity and non-Gaussianity inherent in detector outputs are known to increase false detections of signals. Focusing on identifying the non-Gaussian noise components, this paper introduces a new measure for characterizing the non-Gaussian noise components using the parameter ν which characterizes the weight of tail in a Student-t distribution. A confidence interval is reported on the extent to which detector noise deviates from Gaussianity. Our method revealed stationary and transient deterioration of Gaussianity in LIGO S5 data.

DOI: 10.1103/PhysRevD.93.082005

I. INTRODUCTION

During the last decade, the Laser Interferometer Gravitational-Wave Observatory (LIGO) [1] and Virgo [2] carried out searches for gravitational waves (GWs), and their sensitivities were improved [3–5]. At present these detectors are being upgraded to the advanced LIGO (aLIGO) [6] and the advanced Virgo (AdVirgo) [7] to improve their sensitivities. The Japanese gravitational wave detector, KAGRA [8], is under construction in the Kamioka mine and has cryogenic mirrors whose benefits in detector sensitivity have been verified by a prototype detector, CLIO [9]. The LIGO and Virgo collaborations have set many upper limits on GW amplitudes or event rates of various GW sources [5,10–13], and finally, the aLIGO detected one of the sources, a GW from a binary black hole merger [14,15] in the first observation run. Most of the search methods for GWs are optimal for stationary and Gaussian noise [16]. The sensitivity of interferometric GW detectors is mainly limited by seismic activity, thermal noise, radiation pressure noise, and shot noise in low frequencies above a few tens of hertz. Although these noise components are assumed to be Gaussian in many earlier theoretical studies, it is known from the experiences of the GW searches [17] that detector noise includes nonstationary and/or non-Gaussian noise components. The non-Gaussian components and nonstationary ones make the false alarm rate increase and make it hard to estimate the behavior of noise such as a noise power spectrum. Because these components limit performance of all GW searches, we need to investigate a feature of detector noise for detecting GWs. Reducing false alarms is important for making good use of observational time in the target of opportunity observation for multimessenger astronomy with electromagnetic and neutrino telescopes.

In the past observation runs of LIGO and Virgo, much effort was made to reduce false events by identifying and eliminating narrow-band transient nonstationary noise sources [18–23]. Evaluating the nonstationarity of many auxiliary channels of the GW detectors from the perspective of GW telescope diagnosis improved accuracy of estimating the noise power spectrum and background behavior. The identification and removal of the detector noise and the noise sources has been done by estimating the background behavior in accurately [24,25].

Moreover some methods that are robust against contamination by non-Gaussian noise were suggested [26,27] and showed that the receiver operating characteristic (ROC) curve of a matched filtering method for a GW from a compact binary coalescence from the data of the fifth science run (S5) of LIGO was improved compared with the case of the Gaussian noise model [28,29]. There are also the works that assume non-Gaussian noise in a GW burst search [30].

 © 2016 American Physical Society

Although nonstationarity or non-Gaussianity in GW detector noise were investigated in GW searches and GW telescope diagnosis, sources of these noises are still unknown in many cases. The power of discrimination of nonstationarity and non-Gaussianity must be increased. When non-Gaussian noises exist in detector noise, characteristics of noise such as a power spectrum and a noise event rate can be estimated more accurately by evaluating non-Gaussianity than by regarding all of the deviation from Gaussian noise as nonstationary noise. It is important for dealing with the non-Gaussianity of detector noise in addition to nonstationarity to specify new noise sources.

When a measure for non-Gaussianity insensitive to the stationarity of noise can be defined, such a measure should be helpful to distinguish non-Gaussian and nonstationary noise.

In [28], the ROC curve of the matched filtering method was improved using the Student-t noise model and parameter ν of non-Gaussianity. The Student-t distribution and the parameter ν are defined as

$$p(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right).$$

In order to use this method in the search for GW signals, we must statistically characterize the non-Gaussianity of detector noises. We have to investigate a threshold for distinguishing Gaussian and non-Gaussian noise using ν as a measure. In this paper, the characterization of non-Gaussian noise and our measure of non-Gaussianity are described in Sec. II. In Sec. III, we show the performance of our method at distinguishing non-Gaussian noise from Gaussian noise and statistical assessments of possible misidentification. We demonstrate our method using LIGO data from the fifth science run obtained from the LIGO Open Science Center [31,32]. In Sec. IV, we summarize the performance of our method for finding the degree of non-Gaussianity.

II. NON-GAUSSIAN FEATURE OF DETECTOR NOISE DISTRIBUTION

A. Gaussian noise model

In this paper, we write the time series and Fourier transform of detector noise as $n(t)$ and $\tilde{n}(f)$, respectively. If the time series is Gaussian, the real part $\Re[\tilde{n}(f)]$ and imaginary part $\Im[\tilde{n}(f)]$ of $\tilde{n}(f)$ are Gaussian distributed,

$$p_G(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right), \quad (1)$$

where x is either $\Re[\tilde{n}(f)]$ or $\Im[\tilde{n}(f)]$ and σ the standard deviation of x. An absolute value of $\tilde{n}(f)$, $|\tilde{n}(f)|$, then follows Rayleigh distribution,

$$p_R(x) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right), \quad (2)$$

where x is $|\tilde{n}(f)|$.

B. Student-t noise model

The assumption that the detector noise follows Gaussian distribution often well describes the noise behavior as a zeroth order approximation. However, we cannot always adopt this approximation because there are cases when non-Gaussianity and nonstationarity dominate in a certain time or frequency band. In these cases, the false alarm rate is raised by them.

Figure 1 shows the normalized noise level of LIGO data [31,32] as a function of frequency. The data used were taken from the GPS time 842 747 904 to 842 764 288. This figure shows that the LIGO data are not an ideal Gaussian noise especially in a low-frequency band. This non-Gaussianity in low frequency is probably one of the causes of false alarm events in LIGO S5 observation data [12,23]. Although the non-Gaussian noise is not always dominant, we can find the non-Gaussianity in a certain time and frequency region such as Fig. 1. Investigating such non-Gaussianity may improve the GW search performance compared with the case that regards the detector noise as Gaussian and ignores such non-Gaussianity. So investigating non-Gaussianity enables us to use the detector signal with poor quality.

The Student-t noise model is known as one of the non-Gaussian noise models, which is a natural extension of the Gaussian noise model [33]. In this model $\Re[\tilde{n}(f)]$ and $\Im[\tilde{n}(f)]$ follow the Student-t distribution,

$$p_{\mathrm{ST}}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\sigma_s\Gamma(\frac{\nu}{2})}\left(1+\frac{1}{\nu}\left(\frac{x}{\sigma_s}\right)^2\right)^{-\frac{\nu+1}{2}}, \quad (3)$$

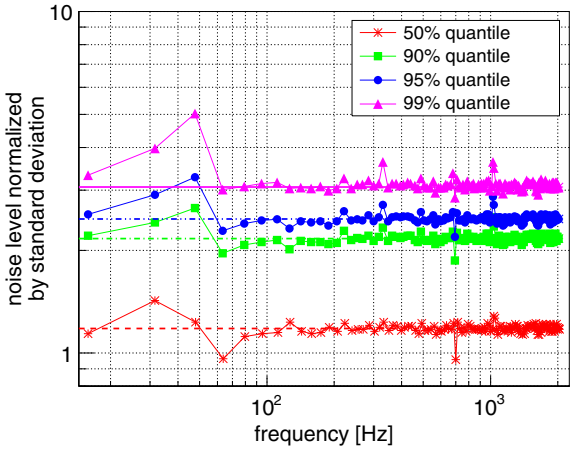

FIG. 1. This plot shows the 50% (red stars), 90% (green squares), 95% (blue circles), and 99% (pink triangles) quantiles of the distribution of the LIGO S5 data taken from GPS time 842 747 904 to 842 764 288. Each quantile was calculated from 2^{18} samples in every 16 Hz. The dashed line, dashed-dotted line, double-dotted line, and triple-dotted line represent the 50%, 90%, 95%, and 99% quantiles expected if the data follow a Gaussian distribution, respectively. This figure indicates that detector noise deviates from Gaussian distribution especially in the low-frequency band.

where x is either $\Re[\tilde{n}(f)]$ or $\Im[\tilde{n}(f)]$, σ_s is a scaling parameter, and ν is a positive value which characterizes the weight of the tail of the distribution. $|\tilde{n}(f)|$ then follows the Student-Rayleigh distribution,

$$p_{\mathrm{SR}}(x) = \frac{x}{\sigma_s^2} p_{F(2,\nu)}\left(\frac{x^2}{2\sigma_s^2}\right), \tag{4}$$

where x is $|\tilde{n}(f)|$ and $p_{F(2,\nu)}$ is the probability density function of the F-distribution with the degrees of freedom $(2,\nu)$ [34,35]. The relation between the scaling parameter, σ_s, and standard deviation of $\Re[\tilde{n}(f)]$ and $\Im[\tilde{n}(f)]$, σ, is $\sigma_s^2 = \frac{\nu-2}{\nu}\sigma^2$. If the variance σ^2 of the detector noise is *a priori* known, some of the statistical properties of our estimator of ν can be calculated analytically. Unfortunately, σ^2 is unknown in general and we must estimate σ^2 from detector noises. So we investigate statistical properties of an estimator of ν with plugging the sample estimate of σ into the expressions. It is noted that Eqs. (3) and (4) are Gaussian and Rayleigh distributions in the $\nu \to \infty$ limit. The Rayleigh and Student-Rayleigh distributions with ν of 4, 8, 16, and 64 are shown in Fig. 2. The weight of the tail is characterized by the parameter ν [36].

It is useful to use ν as a measure which characterizes the deviation from Gaussianity, because the detector noise is known to follow a heavy tail distribution and noise events which belong to the tail of the distribution cause an increase in false alarm events.

It is not necessary to use ν as the measure for non-Gaussianity such that the detector noise always follows the Student-t distribution as long as the detector noise follows a heavy-tailed distribution. In the next section, the method of estimating the ν value of realistic detector noise is described.

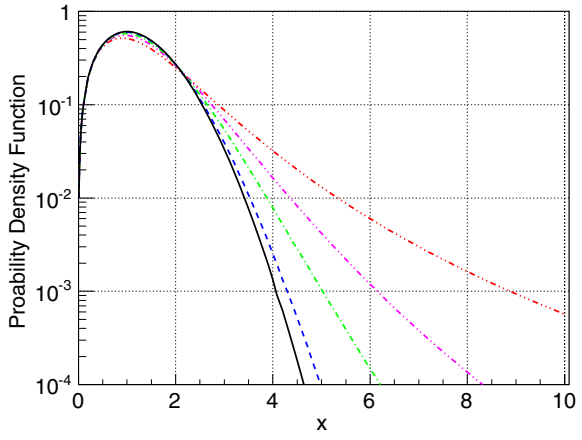

FIG. 2. This plot shows Rayleigh distribution (black solid line) and Student-Rayleigh distribution functions with $\sigma_s = 1$ for various ν. Each color means $\nu = 4$ (blue dashed line), $\nu = 8$ (green dotted line), $\nu = 16$ (pink double-dotted line), and $\nu = 64$ (red triple-dotted line). In the small ν case, the tail of the distribution is heavy. On the other hand, the Student-Rayleigh distribution is closer to the Rayleigh distribution for large ν.

III. NON-GAUSSIAN NOISE CHARACTERIZATION

A. Data processing

An algorithm to estimate ν from the detector noise is shown in Fig. 3. Let us denote the total data length in seconds by T and the sampling frequency in Hz by f_s, and the product of the two by $2M$ (assuming M to be an integer for simplicity). The detector output discrete time series $n(t_k)$ $(k = 1, \ldots, 2M)$ is divided into N chunks $n_i(t)$ with each $t_{FFT} = T/N$ seconds. Then time series data of each chunk $n_i(t)$ is Fourier transformed into $\tilde{n}_i(f_j)(j = 1, \ldots, M)$. We then estimate the parameter ν for the ith chunk at the coarse frequency bin α using l samples around the bin. Specifically, the estimator ν, denoted by $\hat{\nu}_i(f_\alpha)$, is computed from l samples of $\tilde{n}_i(f_j)$ $(j = \alpha - l/2, \ldots, \alpha + l/2)$ using the method explained below. The frequency resolution of $\hat{\nu}_i(f_\alpha)$ is determined by l such that $\delta F = l/t_{\mathrm{FFT}} > 1/T$. In this paper we adopt $t_{\mathrm{FFT}} = 1$ s and $\delta F = 16$ Hz.

The p-quantile, Q_p, is sought from l-sample data around the αth bin in the frequency domain. The estimator $\hat{\nu}$ is defined as

$$\hat{\nu} \equiv \arg\min_{\nu} |Q_p - Q_{\mathrm{SR}}(p,\nu)|, \tag{5}$$

where $Q_{\mathrm{SR}}(p,\nu)$ is the theoretical quantile function of Student-Rayleigh distribution,

$$Q_{\mathrm{SR}}(p,\nu) = \sqrt{2\sigma_s^2 Q_{F(2,\nu)}(p)}, \tag{6}$$

where $Q_{F(2,\nu)}(p)$ is the quantile function of F-distribution with the degrees of freedom (2, ν) [28,34].

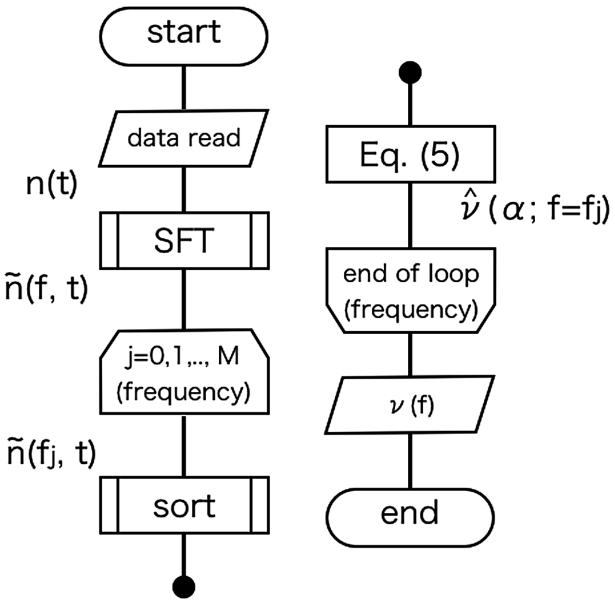

FIG. 3. The algorithm of estimating $\hat{\nu}$ from the detector noise.

We search for ν only in the range from 2.1 to 200 in increments of 0.1 as is shown in Secs. III B and III C. Formally, we need to take the $\nu \to \infty$ limit to recover a Gaussian distribution from a Student-t distribution. In fact, however, the real detector noise can be regarded as virtually Gaussian when $\nu > 200$.

B. Characterization of Gaussian noise

This section shows how $\hat{\nu}$ is used to characterize the non-Gaussianity of data. The performance of $\hat{\nu}$ to quatify non-Gaussianity strongly depends on the data length, T, to be analyzed. We introduce the threshold on ν above which the data follow a Gaussian distribution, taking into account dependency on data length.

524,288 sets of simulated Gaussian noise are generated and analyzed. The length of each data set is 4096 s.

Figure 4 shows the histogram of obtained $\hat{\nu}$. The thick red region represents the 1% lower tail of the $\hat{\nu}$ distribution. If we take the 1% ($\alpha = 0.01$) significance level, the critical region is $\hat{\nu} < \nu_\alpha = 91.4$. Hence, we can reject the null hypothesis that the data follow a Gaussian distribution by 99% confidence if we set the threshold $\nu_{\alpha=0.01} = 91.4$, modulo the error inherited in our Monte Carlo simulations.

When we fix the significance level α, the critical region depends on the data length T because of the sampling variance. Figure 5 shows the threshold $\nu_{\rm th}$ as the function of data length T. The lower regions of the red solid line and blue solid line show the critical regions of rejecting Gaussianity with 99% and 99.9% confidence, respectively.

If the stationarity of data is assumed, the precision of the estimator ν is improved as the data length increases. Since the statistical feature of realistic detector noise varies typically in hours, we need to decide the data length in which the data can be assured to be stationary.

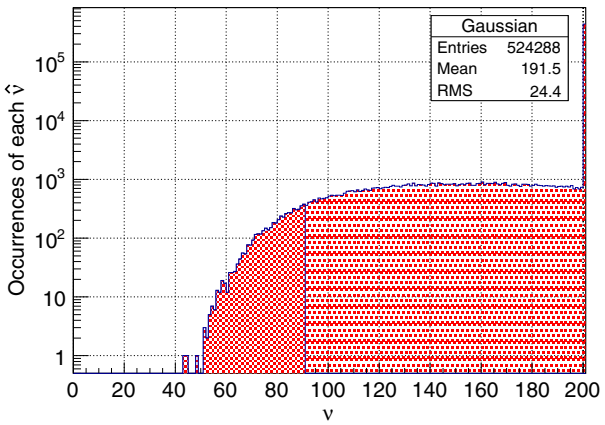

FIG. 4. The resulting histogram of the estimated $\hat{\nu}$ for 524,288 sets of simulated Gaussian noise data. The data length of the simulated noise is 4096 s. This figure shows that even for purely Gaussian noise, $\hat{\nu}$ can be small with a non-negligible probability. The critical region for rejecting Gaussianity is determined by the histogram. When data length $T = 4096$s, $\nu_{\rm th} = 91.4$.

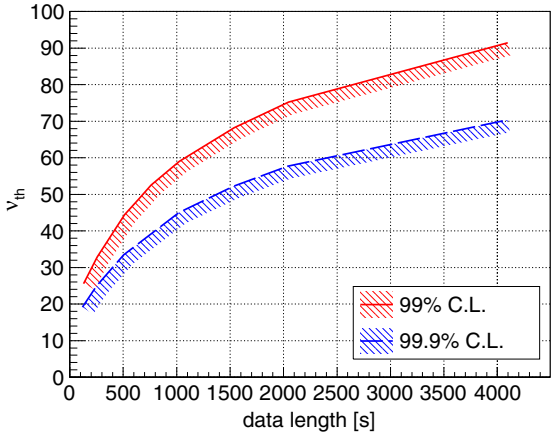

FIG. 5. This plot shows the critical regions for rejecting Gaussianity with significance levels of 1% and 0.1%. In the lower region of each line, Gaussian noise hypothesis is rejected with 99% (red) and 99.9% (blue) confidence. For example, when ν is estimated from 1024 second long data, the hypothesis of Gaussianity is rejected if $\nu \leq 59.0$ with 99% confidence.

Figure 4 shows that $\nu_{\rm th} = 91.4$ in the case of $T = 4096$s with 99% confidence and Fig. 5 shows that the $\nu_{\rm th}$ increases with increasing data length T. Therefore, it is enough that ν is scanned up to 91.4 in the case of $T \leq 4096$s with 99% confidence. However we adopt the range for searching ν as [2.1, 200] throughout this work, because the upper boundary of the range for searching ν, which is larger than $\nu_{\rm th}$, does not affect the conclusion of this paper.

C. Characterization of non-Gaussian noise

In order to evaluate the degree of non-Gaussianity quantitatively it is important to investigate the confidence interval of ν. Our estimator of ν is strongly consistent and therefore asymptotically unbiased.

The confidence interval of ν is difficult to calculate because $\hat{\nu}$ is not asymptotic normal and in general, we do not know the true variance of detector noise. However, the confidence interval can be approximated in the following sense [37].

Proposition: $\hat{\nu}$ is strongly consistent.

Proof: We define $Q_{{\rm SR}\{n\}}(p,\nu)$ as p-quantiles of n samples distributed in a Student-Rayleigh function. By the strong law of large numbers at a given ν,

$$Q^{-1}_{{\rm SR}\{n\}}(Q_{\rm SR}(p,\nu)) \to Q^{-1}_{\rm SR}(Q_{\rm SR}(p,\nu)), \quad \text{almost surely.} \quad (7)$$

Since $p = Q^{-1}_{{\rm SR}\{n\}}(Q_{{\rm SR}\{n\}}(p,\nu)) = Q^{-1}_{\rm SR}(Q_{\rm SR}(p,\nu))$,

$$Q_{{\rm SR}\{n\}}(p,\nu) \to Q_{\rm SR}(p,\nu), \quad \text{almost surely,} \quad (8)$$

by the continuous mapping theorem. $\hat{\nu} \to \nu$ is proved by applying the continuous mapping theorem again because Eq. (6) is also the continuous function of ν. ■

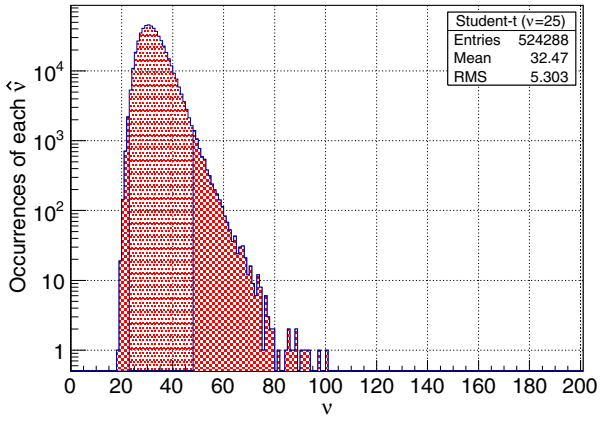

FIG. 6. The result of estimated ν for 524,288 sets of simulated Student-t noise in the case of $\nu = 25$. Estimated ν fluctuate, and the 1% confidence upper ($\nu_{\rm upper}$) and lower ($\nu_{\rm lower}$) boundaries are $\nu_{\rm upper} = 48.8$ and $\nu_{\rm lower} = 23.4$.

Using simulations with a Student-t noise model and various ν, we can find $\nu_{\rm lower}(\nu)$ and $\nu_{\rm upper}(\nu)$ such that $P(\nu_{\rm lower} < \hat{\nu} < \nu_{\rm upper}) = 1 - \alpha$. On the other hand, the confidence interval is $[\hat{\nu}_{\rm lower}, \hat{\nu}_{\rm upper}]$, such that $P(\hat{\nu}_{\rm lower} < \nu < \hat{\nu}_{\rm upper}) = 1 - \alpha$. Here we propose $\hat{\nu}_{\rm lower} := \nu_{\rm lower}(\hat{\nu})$ and $\hat{\nu}_{\rm upper} := \nu_{\rm upper}(\hat{\nu})$. Although there is no rigorous reason why $P(\hat{\nu}_{\rm lower} < \nu < \hat{\nu}_{\rm upper}) = P(\nu_{\rm lower}(\hat{\nu}) < \nu < \nu_{\rm upper}(\hat{\nu}))$ equals $1 - \alpha$, we use $[\nu_{\rm lower}(\hat{\nu}), \nu_{\rm upper}(\hat{\nu})]$ instead of the confidence interval.

$\nu_{\rm upper}$ and $\nu_{\rm lower}$ can be calculated as follows. Figure 6 shows the histogram of the obtained $\hat{\nu}$. The configuration of this simulation is the same as in the previous Gaussian simulation (524,288 sets of simulation noise and data length $T = 4096$ s). The ν of the simulated Student-t noise is 25. The thick red region represents 1% upper $\nu_{\rm upper}$ and lower $\nu_{\rm lower}$ tails of $\hat{\nu}$. The corresponding ν are 23.4 and 48.8, respectively. We performed the same simulations using simulated Student-t noise changing ν from 8 to 50. Figure 7 shows $\nu_{\rm lower}$ and $\nu_{\rm upper}$ as the function of $\hat{\nu}$. In Fig. 7, the red solid line and the blue dashed line represent the lower ($\nu_{\rm lower}$) and upper ($\nu_{\rm upper}$) boundaries of the 98% and 99.8% confidence regions, respectively. The confidence interval is in between $\nu_{\rm lower}$ and $\nu_{\rm upper}$. When $\hat{\nu} = 25$, the 98% confidence interval is $15.45 \leq \nu \leq 27.29$. The confidence interval rapidly spreads for large ν. This deterioration of the confidence interval comes from the fact that the shape of the Student-Rayleigh distribution is more sensitive to the change of ν in the case of small ν than the case of large ν.

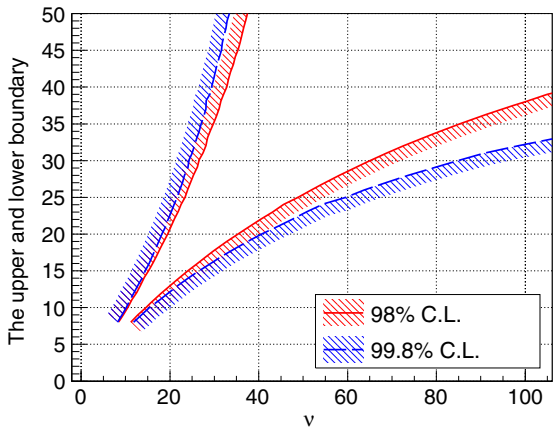

FIG. 7. The 98% (red line) and 99.8% (blue line) confidence interval of ν when the estimation time is 4096 s. For small ν values, resolution of ν is good. For example, when we obtain $\nu = 25$ in the case of 4096 s estimation time, the 98% confidence interval of ν is $15.45 \leq \nu \leq 27.29$.

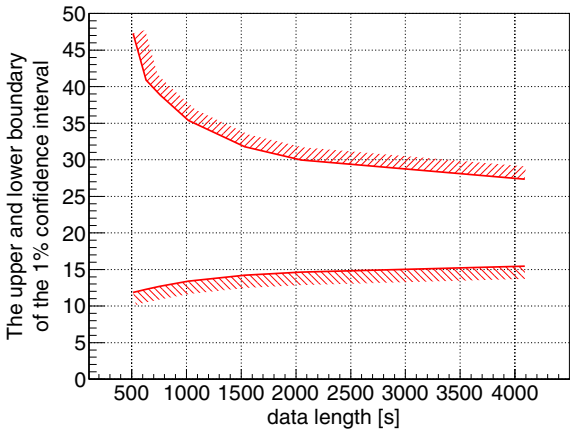

FIG. 8. Dependence of the confidence interval as a function of the data length. The precision of the estimated ν is bad for short data.

Figure 8 shows the dependence of the confidence interval on the data length T. The upper and lower red lines represent $\nu_{\rm upper}$ and $\nu_{\rm lower}$ in the case of Student-t noise with a ν of 25 as a function of T. The confidence interval of $\hat{\nu}$ gets monotonically narrower. When $T = 2048$, 4096, the 98% confidence intervals are $14.52 \leq \nu \leq 29.91$ and $15.45 \leq \nu \leq 27.29$, respectively.

D. Demonstration of our method using LIGO data

We applied our method to LIGO GW strain data from the fifth science run obtained from the LIGO Open Science Center [31,32]. Figure 9 shows the time evolution of the observed $\hat{\nu}$ at each frequency. The GPS time of the data is from 842 747 904 to 842 760 192; the resolution of time δt and frequency δF are 128 s and 16 Hz, respectively. This result is obtained by the data length $T = 1024$ s with overlapped time $T_{\rm lap} = 896$ s.

The $\hat{\nu}$ in the frequency band 30–60 Hz is ~15. The confidence interval is $9.3 \leq \hat{\nu} \leq 15.4$. This non-Gaussianity continues for the whole 16384 seconds when we analyzed the data. The $\hat{\nu}$ in the frequency band 100–1 kHz and the time 8000–9000 s is ~50.

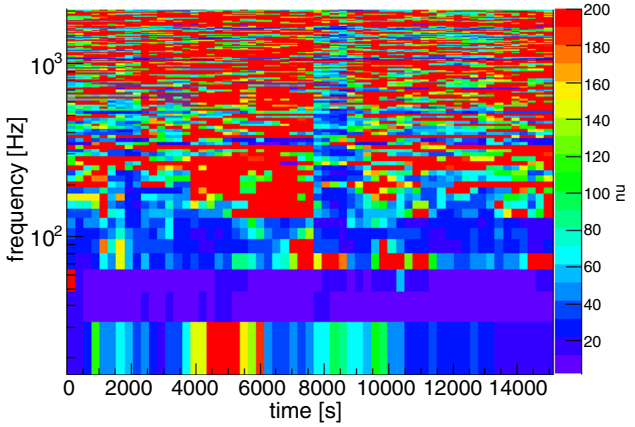

FIG. 9. This plot shows the estimated $\hat{\nu}$ for the real LIGO S5 data as a function of time. The GPS time of the data is from 842 747 904 to 842 760 192; the resolution of time δt and frequency δF are 128 s and 16 Hz, respectively. This result is obtained by the data length $T = 1024$ s with overlapped time $T_{\rm lap} = 896$ s. The purple region means that the Gaussianity of noise is bad and there are many non-Gaussian regions, especially in the low-frequency band.

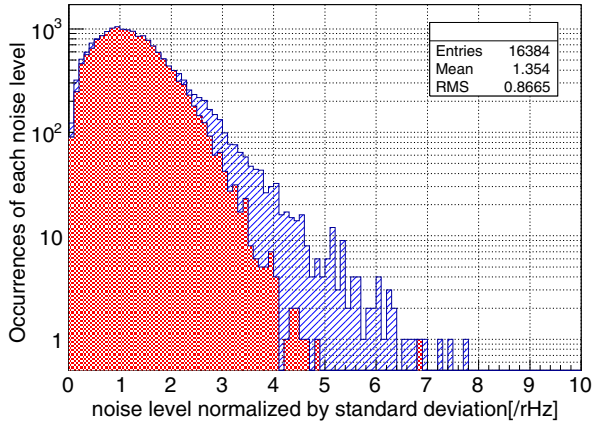

FIG. 10. This plot shows the distributions of the LIGO detector noise in the time-frequency region where time from GPS $= 842\,747\,904$, t; frequency, f; and their widths, dt, df, are ($t = 6144$, $f = 128$, $dt = 128$ s, $df = 16$ Hz) for the red histogram and ($t = 1024$, $f = 32$, $dt = 128$ s, $df = 16$ Hz) for the blue one. The entry of each histogram is 16384.

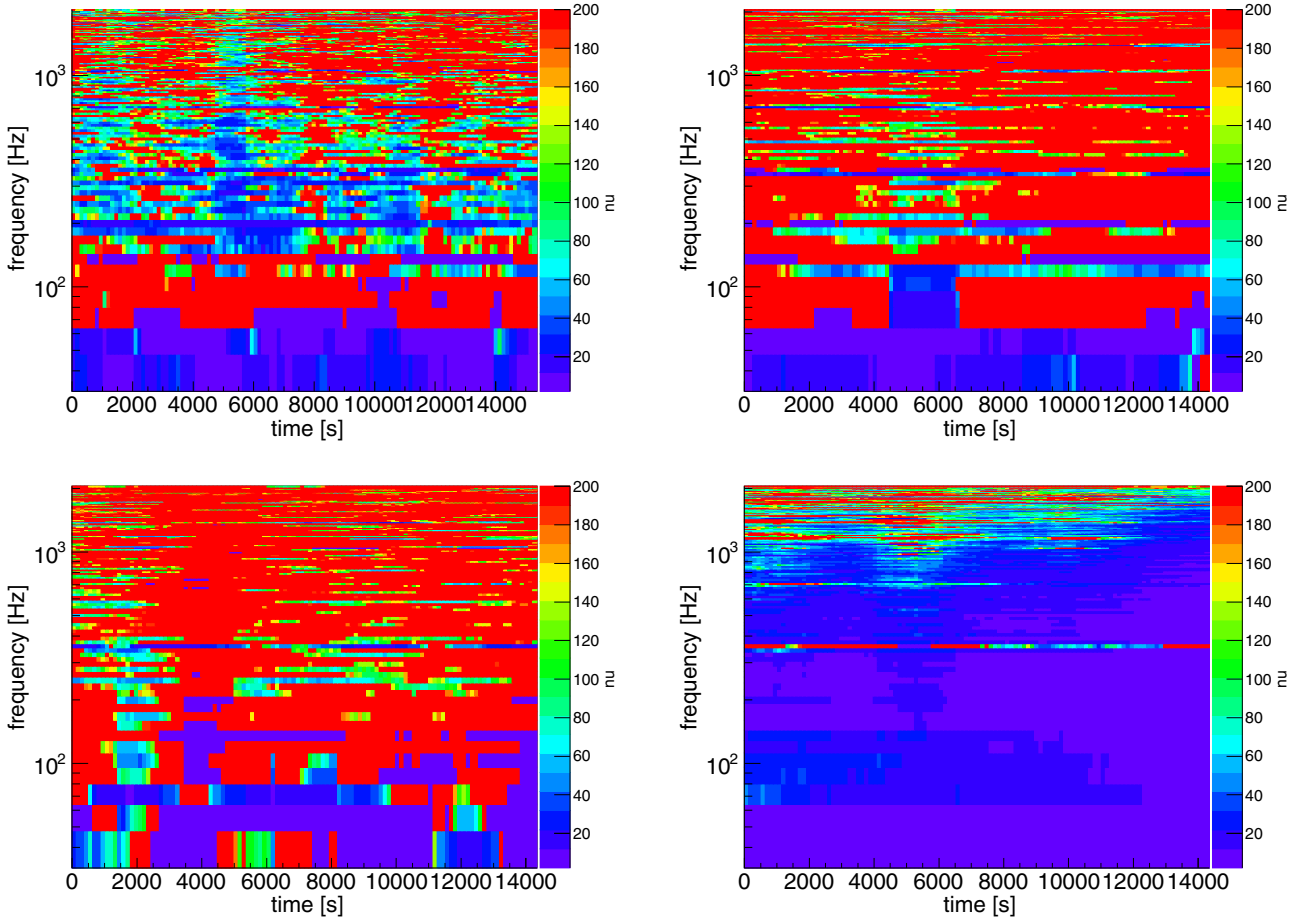

FIG. 11. These plots show the time evolution of the observed $\hat{\nu}$ at each frequency. The GPS times of the top left, top right, bottom left, and bottom right panels are 841 449 472, 842 489 856, 864 575 488, and 870 838 272, respectively. Other parameters are the same as the ones in Fig. 9. The bottom right panel shows the bad quality data which do not belong to CAT4 in the literature of the LIGO and Virgo collaborations [17]. The other panels show good quality data which belong to CAT4.

Data in the regions where we can detect small $\hat{\nu}$ in the spectrogram Fig. 9 are regarded as non-Gaussian by our method (see Fig. 5). This feature is consistent with the quantile plot in Fig. 1. Figure 9 provides us with the temporal evolution of the non-Gaussianity of data. So we can distinguish stationary non-Gaussianity and transient non-Gaussianity.

Deterioration of Gaussianity due to noise transience, which appears around 8000 s, from the 100-to-kHz band is detected. In this region, ν is from 40 to 60 and non-Gaussianity continues for about 300 seconds. From Fig. 5, the assumption that detector noise in this time-frequency region is Gaussian can be rejected with 99% confidence (local probability).

These results show that the detector noise can be distinguished from Gaussian noise quantitatively. In addition to this we succeeded at extracting the degree of non-Gaussianity in the time evolution and frequency band.

Figure 10 shows the distribution of the LIGO detector noise in the time-frequency region where the frequency, the time elapsed from the GPS time = 842 747 904, and the widths of their blue and red histograms are ($f = 32$ Hz, $t = 1024$ s, $df = 16$ Hz, $dt = 128$ s) and ($f = 128$ Hz, $t = 6144$ s, $df = 16$ Hz, $dt = 128$ s), respectively. The estimated $\hat{\nu}$ of the former region is $\sim$200, which can be regarded as Gaussian. That of the latter region is $\sim$15 during the entire time in the figure. The latter region is not Gaussian distributed nor Gaussian with strong outliers caused by glitches. This region seems to originate from noise sources that are stationary and non-Gaussian distributed with heavy tails.

Figure 11 shows the time evolution of the observed $\hat{\nu}$ at each frequency in GPS times different from Fig. 9. The GPS times of the first samples of those data sets are 841 449 472, 842 489 856, 864 575 488, and 870 838 272 with 16384 s of data. Other parameters are the same as the ones in Fig. 9. We have arbitrarily selected those four periods of data except for the selection conditions where those periods (1) do not contain any gap (i.e., no loss of samples) for 18,000 seconds; (2) include no hardware injection; and (3) satisfy CAT1 in the literature of the LIGO and Virgo collaborations [17]. Moreover, we demand that (4) those periods almost satisfy the CAT3/4 conditions except for the one starting at the GPS time 870 838 272, for which we intended to study the effect of CAT4. Note that we demand that condition (1) exclude possible edge effects and use only the middle 16,384 s of data. The deterioration of Gaussianity in the low-frequency band can be revealed in all figures. The bottom right panel shows the observed $\hat{\nu}$ of the bad quality data where data do not belong to CAT4. The Gaussianity of this data can be rejected in almost all frequencies and time with 99% confidence. These figures suggest that the domination of non-Gaussianity appears in the LIGO S5 data even though the data are categorized as CAT4, which represents a good enough quality for science.

IV. CONCLUSION

We proposed a method to characterize the non-Gaussianity of GW detector noise. Namely, we introduced Student-Rayleigh distribution to characterize possible non-Gaussianity of data. Here, the degrees of freedom ν of the distribution are found to be useful to quantify the degree of non-Gaussianity. We used ν as a characteristic parameter that represents the weight of the tail of detector noise. We calculated the confidence interval of ν and the threshold below which the Gaussian hypothesis is rejected. We characterized non-Gaussianity quantitatively in realistic detector noise.

The existence of a non-Gaussian noise component in realistic detector noise was also clarified by our method with a threshold and confidence interval. The threshold for rejecting Gaussian noise was evaluated by changing various data lengths T. In the case of $T = 4096$ s and $\hat{\nu} = 25$, the 99% confidence interval is $15.45 < \nu < 27.29$.

In this work, we showed the detection method for non-Gaussian noise using ν and the accuracy of ν. This method revealed continuous and transient non-Gaussian components in the LIGO data by estimating the degree of non-Gaussianity of the detector noise every 16 Hz and 1024 s.

The degree of non-Gaussianity is related to the origin of the noise source. The characteristics of non-Gaussianity, such as power, frequency, and/or time evolution, are also different among noises of different origins, as shown in the spectrogram Fig. 9. The robustness of outliers derived from the nonstationary noise of our method can be adjusted by changing the quantile and p. The method using quantiles is more robust for outliers than the one using the entire noise distribution, such as χ^2-fitting.

Our method can be used for revealing a stationary feature of the detector noise. So ν can provide different information about the noise status than that provided by the methods for nonstationary noise. When all deviations from Gaussianity are regarded as nonstationarity, it is difficult to identify mechanisms that cause stationary but non-Gaussian noises. Our method and methods for investigating transient noise are complementary to each other for evaluating conditions of the detector noise because our method can investigate stationary non-Gaussianity of detector noise. Regarding the search for GWs, we provide a method to estimate accurate ν that will directly lead to the improvement of the GW search.

ACKNOWLEDGMENTS

This research has made use of data, software, and/or Web tools obtained from the LIGO Open Science Center (https://losc.ligo.org), a service of LIGO Laboratory and the LIGO Scientific Collaboration. LIGO is funded by the United States National Science Foundation. This work is supported in part by the MEXT Grant-in-Aid for the Scientific Research on Innovative Areas "New Developments in

Astrophysics Through Multi-Messenger Observations of Gravitational Wave Sources" (Grant No. 24103005). This work is supported in part by MEXT Leading-edge Research Infrastructure Program, JSPS Grant-in-Aid for Specially Promoted Research No. 26000005. We thank S. D. Mohanty for valuable discussions and helpful comments. K. H. would like to thank B. Allen for warm hospitality during his stay in Hannover. K. H. is thankful to M.-K. Fujimoto for continuing encouragement. This work is supported by the JSPS Grants-in-Aid for Scientific Research (KAKENHI) 25800126 and 15K05070 (Y. I.).

[1] B. P. Abbott *et al.*, Rep. Prog. Phys. **72**, 076901 (2009).
[2] T. Accadia *et al.*, J. Instrum. **7**, P03012 (2012).
[3] B. Abbott *et al.*, Phys. Rev. D **77**, 062004 (2008).
[4] J. Abadie *et al.*, Astrophys. J. **734**, L35 (2011).
[5] J. Abadie *et al.*, Phys. Rev. D **85**, 122007 (2012).
[6] G. M. Harry (LIGO Scientific Collaboration), Classical Quantum Gravity **27**, 084006 (2010).
[7] F. Acernese *et al.*, Classical Quantum Gravity **32**, 024001 (2015).
[8] K. Kuroda *et al.*, Classical Quantum Gravity **27**, 084004 (2010).
[9] M. Ohashi *et al.*, Classical Quantum Gravity **20**, S599 (2003).
[10] J. Aasi *et al.*, Phys. Rev. D **89**, 122003 (2014).
[11] J. Aasi *et al.*, Phys. Rev. Lett. **113**, 231101 (2014).
[12] J. Aasi *et al.*, Phys. Rev. D **91**, 062008 (2015).
[13] M. A. Bizouard and M. A. Papa, C.R. Phys. **14**, 352 (2013).
[14] B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 061102 (2016).
[15] B. P. Abbott *et al.*, Astrophys. J. Lett. **818**, L22 (2016).
[16] P. Jaranowski and A. Królak, Living Rev. Relativity **15**, 4 (2012).
[17] J. Aasi *et al.*, Classical Quantum Gravity **29**, 155002 (2012).
[18] Jessica McIver, Classical Quantum Gravity **29**, 124010 (2012).
[19] J. R Smith, T. Abbott, E. Hirose, N. Leroy, D. MacLeod, J. McIver, P. Saulson, and P. Shawhan, Classical Quantum Gravity **28**, 235005 (2011).
[20] P. Ajith, T. Isogai, N. Christensen, R. X. Adhikari, A. B. Pearlman, A. Wein, A. J. Weinstein, and B. Yuan, Phys. Rev. D **89**, 122001 (2014).
[21] N. Christensen (LIGO and VIRGO Scientific Collaboration), Classical Quantum Gravity **27**, 194010 (2010).
[22] J. Slutsky *et al.*, Classical Quantum Gravity **27**, 165023 (2010).
[23] J. Aasi *et al.*, Classical Quantum Gravity **32**, 115012 (2015).
[24] M. Coughlin, J. Phys. Conf. Ser. **243**, 012010 (2010).
[25] M. W. Coughlin, Classical Quantum Gravity **28**, 235008 (2011).
[26] J. D. Creighton, Phys. Rev. D **60**, 021101 (1999).
[27] B. Allen, J. D. Creighton, E. E. Flanagan, and J. D. Romano, Phys. Rev. D **65**, 122002 (2002).
[28] C. Röver, Phys. Rev. D **84**, 122004 (2011).
[29] R. Biswas *et al.*, Phys. Rev. D **85**, 122009 (2012).
[30] M. Principe and I. M. Pinto, Classical Quantum Gravity **26**, 204001 (2009).
[31] LIGO Scientific Collaboration, LIGO Open Science Center release of S5, doi:10.7935/K5WD3XHR (2014).
[32] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, J. Phys. Conf. Ser. **610**, 012021 (2015).
[33] Student, Biometrika **6**, 1 (1908).
[34] C. Röver, R. Meyer, and N. Christensen, Classical Quantum Gravity **28**, 015010 (2011).
[35] C. Röver, LIGO Document, Report No. LIGO-T1100497-v1, https://dcc.ligo.org/LIGO-T1100497/public.
[36] K. L. Lange *et al.*, J. Am. Stat. Assoc. **84**, 881 (1989).
[37] J. E. Gentle, *Elements of Computational Statistics* (Springer-Verlag, New York, 2002).